Rahul Duggal[1], Hao Zhou[2], Shuo Yang[2], Jun Fang[2], Yuanjun Xiong[2], Wei Xia[2]

[1]Georgia Tech, [2]AWS AI

Abstract: This work tackles the emergent problem of reducing negative flips: test samples that are correctly predicted by a less accurate model, but incorrectly by a more accurate one. We introduce REG-NAS to design a family of highly accurate models that engender fewer negative flips. It consists of two components:

1. A novel architecture constraint that enables a larger model to contain all the weights of the smaller one thus maximizing weight sharing.
2. A novel search reward that incorporates both Top-1 accuracy and negative flips in the architecture search metric.

We demonstrate that REG-NAS can successfully find desirable architectures with 33-48% fewer negative flips in three popular architecture search spaces.
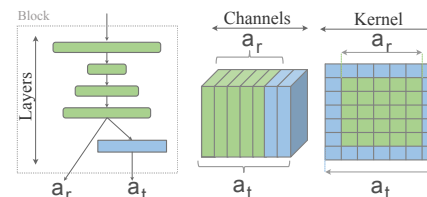
Designing Regression-Free Models
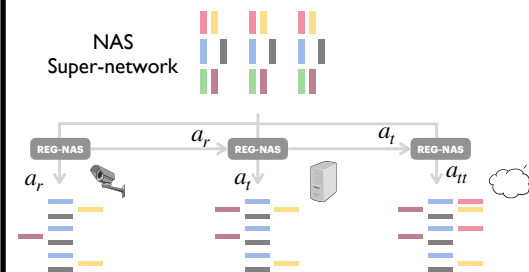
$$\text{Reward} = \lambda_1 * Top1(a_t) - \lambda_2 * NFR(a_t, a_r)$$

$$\mathcal{R}_0 : \lambda_1 = 1, \lambda_2 = 0$$
$$\mathcal{R}_1 : \lambda_1 = 0, \lambda_2 = 1$$
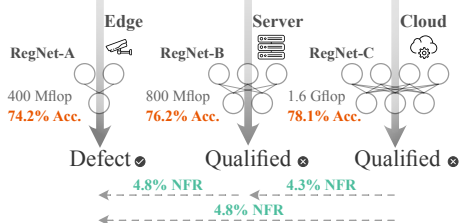$$\mathcal{R}_2 : \lambda_1 = 1, \lambda_2 = 1$$

(a) Architecture constraint   (b) Search reward

Fig. 2 REG-NAS workflow.

Fig. 3 Two components of REG-NAS.

Key Results

Fig. 6 Comparing model design strategies.

Fig. 7 Properties of REG-NAS.

Fig. 8 Generalization to other search spaces.

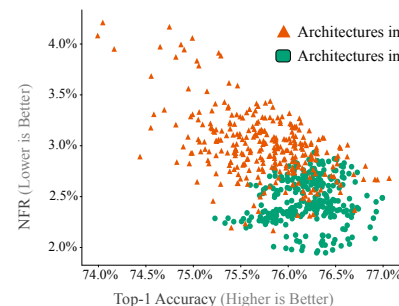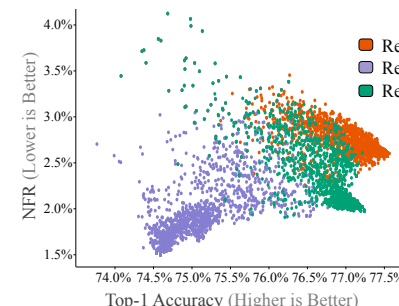## AI on Diverse Compute Platforms
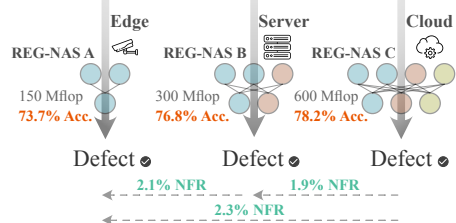
Fig. 1 Illustrating the challenge of negative flips.

Effect of REG-NAS

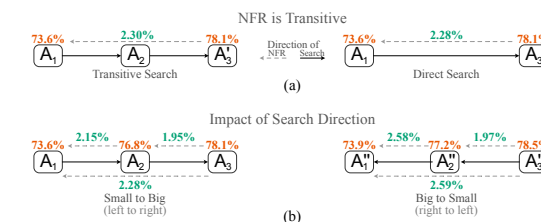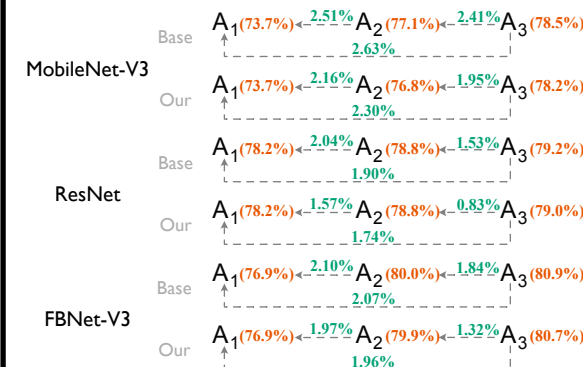**Fig. 4** Effect of the architecture constraint.   **Fig. 5** Effect of the search reward.